Guidelines for the creation of a data management plan

Definition and purpose

A data management plan (DMP) is a document that describes the future handling of research data. The plan includes documentation of the data, their archiving and long-term availability after the project has been completed. A DMP clarifies who holds which responsibilities and obligations in regards to the data. The advantage of creating a DMP at the beginning of the project is that challenges can be identified at an early stage and suitable strategies can be developed to ensure the reusability of the data and make the research process more efficient. Data management plans are living documents that can be updated and fleshed out as the project progresses.

Specifications and support

Many research funders and institutions require the creation of a DMP and have specific requirements for handling research data. Please find out about the requirements and discipline-specific features that apply to you.

Notable documents are TUM Guidelines for Handling Research Data, the Bylaw of Good Scientific Practice which in §8 deals with the protection and storage of primary data, as well as the TUM Open Access Policy.

Support for writing a DMP is provided by, among others:

- The German Research Foundation's checklist for handling research data
- The Horizon Europe DMP Template of the European Commission
- DMP tools, such as RDMO, DS-Wizard, DMPonline, and DMPTool
- Subject-specific comments in the portal of the DFG Code of Conduct "Guidelines for Safeguarding Good Scientific Practice"
- Services and recommendations of the NFDI consortia

This cross-disciplinary handout from the University Library addresses the key components of a DMP and contains tips and examples to help researchers carefully plan handling of research data and fulfill the FAIR principles (data should be findable, accessible, interoperable, reusable). For further assistance on the individual aspects, please contact TUM Research Data Hub.

Components of the data management plan

1. Administrative Information

- What is your research project about? State the title, duration, aim, framework conditions and context of the project.

- Who is responsible for research data management throughout the project? Specify who is responsible for tasks such as documentation, quality control, archiving and publication of the data.

- Which guidelines and policies must be observed in the project? Check the guidelines of TUM and, if applicable, the research funding organization. Will a project-specific policy be drawn up if necessary?

2. Description of the data

Data collection
- Are existing data being reused? Check whether and to what extent reuse of the data is permitted and cite the data source.

- What different types of data (e.g. measurement data, laboratory values, surveys, observation data) are generated in the project?

- How are the data collected and processed? Are devices or specific software used to collect or analyze the data?

- In which format are the data available? When choosing the format, use subject-specific standards as a guide. Simple, non-proprietary and widely used formats are particularly suitable for long-term storage of digital research data (see the University Library's list of recommended data formats).

- How much data do you expect in the project? Even an approximate estimate is helpful in identifying suitable storage locations and likely costs.

Documentation
- How are methods and procedures for collecting and processing data documented? For example, do you use lab books, code books, README files, log files, annotations within the files?

- Is there a standardized metadata collection? Also consider existing metadata standards in your subject area and the interdisciplinary nature of the selected metadata. You can find lists of metadata standards here: Disciplinary Metadata, Metadata Standards Catalog, NFDI.

- Are your data reproducible? Describe which requirements (e.g. equipment, effort, costs) are necessary for the reproducibility of the research results.

- How do you structure and name your files? Are the files organized according to sub-projects, methods, or specific research questions? Is there a convention for file naming, e.g. YYYYMMDD_subject_method_version?

- How do you handle different file versions? For example, do you create manual copies or do you use version control software such as Git? How are the differences between versions documented?

Quality control
- How is the quality of the data ensured? For example, are experiments repeated, cross-checks made with other data or automatic plausibility checks carried out?

- How is the quality of the metadata ensured? Who assigns or checks the metadata? Is the metadata generated (semi-)automatically, e.g. by devices or metadata tools?

3. Data backup

- Where are the data stored during the project? Are all the project data stored centrally in one place or are they stored locally by the project partners and only exchanged when necessary? The Leibniz Supercomputing Centre (LRZ) offers storage solutions such as personal cloud storage, project cloud storage for working groups, Data Science Storage for large amounts of data and Sync+Share for collaborative work.

- How do you allow project members to access the data? Are authorizations and roles assigned? Can access to data be granted to external project partners?

- How do you handle confidential data? Are secure authentication methods, strong passwords, and encryption measures in place?

- What backup strategies do you use? The LRZ's archive backup system, based on IBM Spectrum Protect software, enables regular and automated data backups of servers and computers.

- What data are archived long-term after the project is completed? The DFG recommends a retention period of 10 years. Depending on the discipline, type of data, and project-specific requirements, other retention periods may apply.

- If your data is not available in digital form, you must also state where and for how long the data will be stored.

4. Data publication

- What is the potential for reuse of your data? Describe the benefits of the data for third parties. In particular, consider which data is impossible, difficult, or expensive to reproduce and should therefore be published.

- Where would you like to publish your data? Over 2000 repositories for research data are listed in the re3data register. Give reasons for choosing a repository, e.g. costs, allocation of persistent identifiers, licenses, dissemination in the community. In the institutional media server mediaTUM, data can be published free of charge, licenses can be issued, and DOIs (Digital Object Identifiers) as well as concept DOIs for versioning can be registered.

- Which licenses are granted? If the research data reaches the level of copyright, the data can licensed. The licenses determine the extent to which the research data may be reused. Unless there are reasons not to do so, open licenses should be chosen. They facilitate the reuse of the data in the spirit of the Open Access movement. Further information on the choice of licenses can be found in the handout for the publication of scientific data and software programs of the University Library.

- Are certain research data not published or published with restrictions? Indicate whether and how access to the data is restricted or whether the data is only published after an embargo period has expired. Give reasons for this, e.g. protection of personal data or confidential company data, patent claims, etc.

5. Legal and ethical aspects

- Is copyright observed when data is reused and licenses are granted? Check the extent to which data may be processed and redistributed. Also consider contractual agreements between you and your employer, the research funder, and your project partners. Further information can also be found in the handout on legal aspects of research data management of the University Library.

- Are personal data collected? Describe the measures you take to comply with the legal requirements for data protection and data security. For example, are the data subjects' declaration of consent obtained? Is personal data anonymized or pseudonymized? If you have any data protection questions, please contact the data protection officer at your institution.

- Are there any ethical aspects to consider in your project? Explain whether the data collection could have negative health, social, or economic effects on study participants. Check whether an ethics vote must be obtained from the TUM Ethics Committee.

6. Cost estimate

What costs are incurred for data management? List the expected costs and find out about funding opportunities. For example, you could apply for funding for the following costs:

- Software/hardware, e.g. costs for the collection and processing of data or for the use of existing infrastructures

- Data publication, e.g. costs depending on the choice of repository, data size, and possibly other services such as data curation, quality checks, and peer review procedures

- Personnel, e.g. costs for the position of a data steward who is responsible for research data management in the project. TUM Research Data Hub will be happy to inform you about TUM's support options for filling data steward positions.

Technical University of Munich
University Library
Research Data Services

researchdata@tum.de
www.ub.tum.de
www.tum.de

Last modified: June 2024